DISTRIBUTED RETRIEVAL PRACTICE ENHANCES PRIMARY SCHOOL STUDENTS' RETENTION OF COMPUTATIONAL THINKING CONCEPTS

Lydia Casanova, Emily Theophilou, Marc Beardsley, Patricia Santos and Davinia Hernández-Leo
Universitat Pompeu Fabra – Barcelona, Spain, 08018

ABSTRACT

Efforts are being made to add Computational Thinking (CT) to the curriculum of primary schools in order to develop student abilities to use computational tools to solve problems and better prepare them for an increasingly digital society. Distributed Practice (DP) and Retrieval Practice (RP) are evidence-based practices that have shown to lead to efficient and durable learning. Studies investigating Distributed Retrieval Practice (DRP), a combination of the two evidence-based practices, have found positive effects of DRP on university student learning of anatomy and vocabulary items. As CT is considered to be a complex thinking skill, one that is acquired over time with practice, we investigated whether DRP had durable effects on primary student learning of CT concepts. We conducted a quasi-experimental study involving 20 primary school students between 11 and 12 years of age. The students participated in 6 weekly sessions on CT and completed two tests – one at the end of the fifth session (immediate) and the other a week later (delayed). At the end of each of the first 4 sessions, students performed DRP by completing a class-wide review quiz consisting of 5 multiple-choice questions. The results of the study show that students attained higher scores on the delayed test (72.6% ± 19.2) in comparison to the immediate test (67.9% ± 20.7). However, the difference in scores was not significant. The results suggest that DRP may play a role in helping students overcome the natural decay of memory, however, studies with a larger sample size are required. Further, this study demonstrates that DRP can be applied in an authentic classroom environment that involves primary school student learning of a complex subject such as CT.

KEYWORDS

Distributed Retrieval Practice, Computational Thinking, Distributed Practice, Retrieval Practice

1. INTRODUCTION

Computational Thinking (CT) is shorthand for thinking like a computer scientist. In other words, it is "the ability to use the concepts of computer science to formulate and solve problems" (European Commission, 2019) and involves logical thinking, pattern recognition, and decomposition skills. Recent initiatives such as (EU) CompuThink, designed and funded by the Joint Research Centre of the European Commission, aim to introduce CT to primary and secondary school students (Bocconi et al., 2016). A key objective behind teaching CT is to develop student abilities to use computational tools to solve problems and better prepare students for an increasingly digital society (Yadav et al., 2016). To fulfill the objective of it being a useful skill for the future, students need to develop and retain a level of proficiency in CT. As it is considered to be a complex thinking skill, that is acquired over time with practice (Denning, 2017), appropriate teaching practices need to be explored to allow the acquired knowledge and skill to persist over time. However, learning can quickly become forgotten as demonstrated by the work of Ebbinghaus almost 135 years ago (Murre and Dros, 2015). With the rapid loss of memory over time, long-term or durable learning is hindered (Davis and Zhong, 2017).

Distributed Practice (DP) and Retrieval Practice (RP) are two evidence-based practices that have been shown to lead to efficient and durable learning. DP consists of distributing or spacing reviews over time instead of massing them in a single session (Murray and Udermann, 2003). RP consists of recalling information from memory instead of relying on information in hand. It is a strategy that spurs the learner to produce an effort from within to induce better retention (Roediger & Butler, 2011). Distributed Retrieval Practice (DRP) is a combination of the two evidence-based practices. Studies investigating DRP have found positive effects of DRP on university student learning of anatomy (Dobson et al. 2017) and vocabulary items (Nakata et al., 2020).

In our study, we looked to explore the effects of DRP on primary school student learning of a complex subject such as CT in an authentic classroom setting.

This paper is structured in the following manner. First, we present a review of literature related to the concepts that motivate our research question and support our hypothesis. This is followed by a description of the methods used for our study and the results. Finally, we discuss the results, limitations of the study, and present a conclusion of the overall work.

2. STATE OF THE ART

2.1 Durable Learning

An objective of formal education is for students to be able to apply the skills and knowledge acquired in the classroom to future contexts outside of the classroom. To achieve this, student learning needs to be durable and remain beyond immediate assessments (Halpern and Hakel, 2002). In writing about learning, Davis and Zhong (2017) explain that the process of learning involves the creation of memories. Memories are created by a symphony of different neurons that work together to create, store, and later allow us to retrieve information. Within this process, there is an ongoing balance between the strengthening forces of acquisition and consolidation and the weakening forces of forgetting in forming a memory. These weakening forces are referred to as the process of active forgetting which, along with acquisition and consolidation, are parts of the brain's biological system for managing memories. In other words, active forgetting contributes to the eroding of memory traces. These forces of forgetting were demonstrated in Ebbinghaus' seminal work on the forgetting curve almost 135 years ago. Work that has recently been reproduced by Murre and Dros (2015) shows that on average we remember as little as 30% of what we learned after one day has passed.

Research on evidence-based learning practices also demonstrates that much of what we learn is forgotten, unless appropriate learning practices are used to strengthen the memories of what we have learned. A study in 2006 (Roediger et al., 2006), investigated whether testing promotes longer retention than rereading in a comparison of study strategies. Subjects in both the testing and rereading conditions completed an immediate (same day) and delayed test (1 week later). The results of the study showed that both conditions performed worse on the delayed test in comparison to the immediate test. However, subjects in the testing condition only saw their performance decrease by 14%, whereas participants in the rereading condition saw a decrease of 52%. The study demonstrates that the retrieval of information can benefit memory retention and also provides an example of how much forgetting occurs over time. Interestingly the results of some studies suggest that DP can help learners circumvent the natural decay of memories. For example, the results of a study conducted by Rawson and Kintsch (2005) showed that students reviewing the target material following a distributed pattern performed better on the delayed test in comparison to an immediate test.

2.2 Distributed Retrieval Practice

Studies investigating Distributed Retrieval Practice (DRP), a combination of the two evidence-based practices, have found positive effects of DRP on student learning. For example, a study involving university students' learning anatomy information investigated whether the benefits of RP varied between massed and distributed learning conditions. Study results revealed that distributed strategies produced a significantly better recall of anatomy information than massed strategies (Dobson et al. 2017). Another study involving university students' learning vocabulary investigated the effects of cumulative and noncumulative quizzes taken across 9 weeks. The study found that cumulative tests (i.e. DP) are as much as three times more effective than noncumulative tests when comparing the results of a delayed test taken 3 weeks later (Nakata et al., 2020).

This leads to our research question: Does DRP positively affect primary student learning of a complex subject such as CT? Based on current research related to DP, RP, and DRP, we hypothesize that DRP will positively affect primary student learning of CT with students scoring higher on a delayed test of CT concepts in comparison to an immediate test.

3. METHODS

3.1 Study Context and Participants

The present study was conducted in the context of the project "Makers a les Aules" (Makers in the Classroom), an initiative carried out by a public university (providing the instructors of the sessions) and the Barcelona City Council. The goal of the project was to support primary school teachers in introducing maker activities to their students. Participating schools were public primary schools from Barcelona, Spain. In total 10 schools participated in the project with one group of students per school completing the Makers a les Aules activities across 5-6 sessions. Of the 10 groups of students, three with similar profiles were selected to participate in the study. However, due to the COVID-19 pandemic, only one of the groups (11 to 12-year-olds) was able to complete the study. The study was conducted in Catalan.

3.2 Experimental Design and Materials

We designed a quasi-experimental study that took place in an authentic classroom setting. The participating students attended 6 weekly sessions on CT (see Table 1) and completed two tests – one at the end of the fifth session (immediate) and the other a week later (delayed). Further, at the end of each of the first 4 sessions, students performed DRP by completing a class-wide review quiz consisting of 5 multiple-choice questions. A more detailed explanation of the experimental design and materials follows.

For the study, the following approach was used to teach students about CT via maker activities (Papavlasopoulou et al., 2017). Students participated in 6 weekly sessions. The first 5 sessions were learning sessions in which students were taught CT concepts and applied them toward creating a working prototype of an interactive mockup using Makey Makey (Silver et al., 2012). The final sixth session was a demonstration session in which groups of students showed demos of their prototypes. The 90-minute learning sessions taught programming with Scratch, as programming has been found to be an effective way to teach CT (Rich et al., 2013). Scratch is a simplified coding environment in which learners drag visual coding blocks in a canvas to control characters, images, sounds, and text to create different animations, games, stories, and art (Resnick et al., 2009). Despite having a simplified coding environment, Scratch still allows students to become familiar with the logic and structures involved in programming (Kelleher and Pausch, 2005). Table 2 shows the design of each classroom session. In the first learning session, the opening 20 minutes were spent on a pre-questionnaire, 15 minutes on theoretical content, 40 minutes for hands-on practice, and 15 minutes for a session review. In learning sessions 2-4, 15 minutes were allocated to theoretical content delivered at the beginning of the session, 60 minutes for practical hands-on applications such as Scratch programming or interconnecting elements using Makey Makey, and 15 minutes for a session review. In learning session 5, the only differences to the previous learning sessions were that the hands-on practice was 5 minutes shorter and an immediate test was given to students during the final 20 minutes of the session instead of the session review. The theoretical content which covered basic concepts from CT such as identification and understanding of loops, conditions, inputs, and outputs (see Table 1 for all topics) were introduced by an instructor from the university, using visual presentation slides combined with oral explanations. For the demonstration session no new material was introduced. The session was 60 minutes with the first 40 minutes spent on group demos. During the final 20 minutes students took the delayed test.

To apply DRP, students participated in a class-wide review quiz during the final 15 minutes of each of the first 4 sessions. The quiz provided an opportunity for RP. Further, the quiz questions were drawn from both the immediate and preceding sessions which permitted DP. A previous study applied DP (Rohrer, 2009) to a mathematics course by including a daily assessment for students with exercises from different chapters covered in earlier lessons mixed in. This approach improved the long-term performance of students. We followed a similar approach by mixing up the questions used in the 5-question, multiple-choice class review quiz. Table 1 shows how the questions were distributed per session. Kahoot!, a free online software that allows teachers to create quizzes and share them with students creating a contest-like environment, was used to deliver the review quiz to the class. Studies have shown that Kahoot! can have a greater positive effect on learning when compared to other tools and approaches (Wang and Tahir, 2020). Studies have also shown that receiving prompt feedback after RP can help improve student learning by allowing learners to reflect on their given answers and improve

on them (Butler and Roediger, 2008; Pashler et al., 2005). Accordingly, after each review quiz question, students received immediate feedback on their answers.

Two tests were used to assess student learning of CT concepts. One test was given at the end of the final fifth session (immediate test) and the other test was given a week later (delayed test). Both tests contained 10 true or false questions (2 questions from each session) that needed to be answered within 20 minutes. Each question was presented with 3 possible answers: ‘True’, ‘False’, and ‘I don’t know’. These answers were scored as +1, -1, and 0 respectively. This formula scoring method gives students the opportunity to acknowledge that they do not know the correct answer instead of forcing them to guess (Muijtjens et. al 1999). Formula scoring offers an individualized way of correction for guessing and may reduce random guessing to as low as 2% of the items (Cecilio-Fernandes et al 2017). It should be noted that students were not warned in advance about the tests in an effort to reduce the likelihood of students reviewing the materials prior to the tests.

Note that in session 1, students also completed a 5-question pre-questionnaire. The pre-questionnaire explored student familiarity and prior knowledge related to CT, Scratch, and Makey Makey. The questionnaire consisted of two sections. The first section asked students about their prior experience with Scratch and Makey Makey. It consisted of binary answers which were quantified in order to average students’ overall experience with the tools (Yes=1, No= 0). The second section aimed to gauge student competence related to CT. It consisted of pseudo-code and an exercise related to it. This question was scored as 1 if answered correctly or 0 answered incorrectly.

Table 1. Overview of classroom sessions

DRP	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6
Sessions	S1	S2	S3	S4	S5	S6
Concepts	Introduction to Makey Makey Algorithm definition	Inequality symbols Variables definition	Identification and understanding of conditions	Identification and understanding of loops	Identification and understanding of inputs and outputs	No content provided Demonstration of prototypes
Experimental activities	Pre-questionnaire 5 questions S1	3 questions S2 2 questions S1	3 questions S3 2 questions S2	3 questions S4 1 question S3 1 question S1	Immediate test	Delayed test

Table 2. Design of the classroom sessions

Learning sessions	Session 1	Session 2-4	Session 5	Session 6
Duration	90 minutes	90 minutes	90 minutes	80 minutes
Time distribution	Pre-questionnaire [20 minutes] Theoretical content [15 minutes] Hands-on practice [40 minutes] Session review [15 minutes]	Theoretical content [15 minutes] Hands-on practice [60 minutes] Session review [15 minutes]	Theoretical content [15 minutes] Hands-on practice [55 minutes] Immediate test [20 minutes]	Student demos [60 minutes] Delayed test [20 minutes]

4. RESULTS

A total of 20 out of 25 eligible students completed the study. Five participants needed to be removed from the study as they did not complete both the immediate and delayed tests. Participants (11 boys, 9 girls) were between the ages of 11 and 12 ($M = 11.15$, $SD = 0.37$). Further, the results of the pre-questionnaire revealed that all students had previous knowledge of Scratch, few were familiar with Makey Makey (4.0 %), and 88% of the students answered the CT-related problem correctly.

4.1 Immediate and Delayed Test

To determine whether students scored higher on a delayed test of CT concepts in comparison to an immediate test, we conducted a dependent t-test. Firstly, the Kolmogorov-Smirnov Test for normality was conducted and results supported our choice of the next statistical measure. Subsequently, a dependent t-test was conducted with the following null hypothesis: average scores from immediate and delayed tests are the same. Students attained higher scores on the delayed test ($M = 72.6$, $SD = 19.2$) in comparison to the immediate test ($M = 67.9$, $SD = 20.7$). However, the difference between the immediate and delayed test scores was not significant ($t(18) = -0.91$, $p = 0.19$). The distribution of scores and statistics from the immediate and delayed tests can be seen in Figure 1 and Table 2.

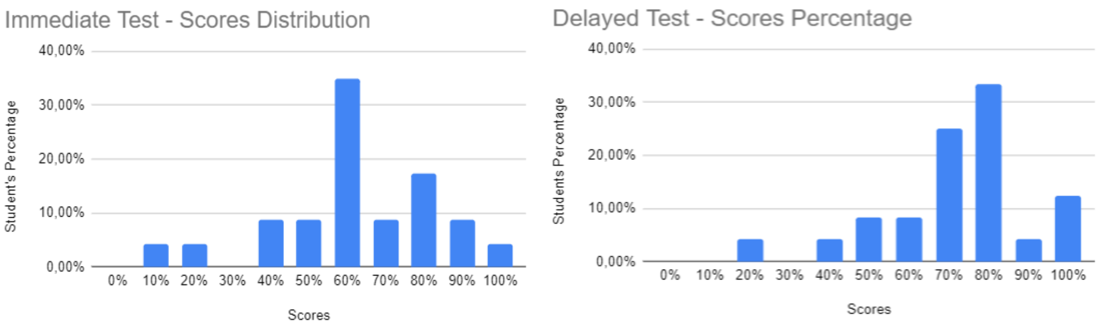

Figure 1. Distributions of Immediate and Delayed Test Scores

Table 2. Analysis of Immediate and Delayed Test Scores

Statistics	Immediate Test	Delayed Test
Average Score	67.90%	72.60%
Standard Deviation	20.7	19.2
Variance	450.9	387.1
Min score	20%	20%
Max score	100%	100%
% Fails (<50%)	15%	5%
% Pass (50% or 60%)	35%	20%
% Remarkable (70% or 80%)	30%	55%
% Excellent (90% or 100%)	20%	20%

5. DISCUSSION

5.1 Interpretation and Implication of Results

We hypothesized that students would attain significantly better results on the delayed test when compared to the immediate one. However, the p-value of the t-test showed that there was no significant difference between the two tests. Despite not obtaining significantly higher scores on the delayed test, the results of the delayed scores were not worse than the immediate scores suggesting that DRP helps learners overcome the natural decay of memory and this is consistent with Rawson and Kintsch´s study in which the performance of students following a DP pattern improved and did not worsen on a delayed test (Rawson & Kintsch, 2005). Nevertheless, our results are inconsistent with the study by Roediger et al. (2006) in which the performance of students under the RP condition performed worse on the delayed test compared to the immediate test. Thus, the enhanced results on the delayed test might be related to DP rather than RP. According to our results, DRP seems to be an educational method that could improve long-term memory retention and not hinder it. As DRP can be easily applied at the end of each session (only 15 minutes were used in this case) to review past content, it is worth exploring further to see if it can help the knowledge and skills students acquire in class to persist over longer durations.

Studies have reported that prior knowledge is a key factor that must be considered in order to evaluate the learning process (Pazzani, 1991) and to control for prior knowledge as a possible confound. Accordingly, we had students complete a pre-questionnaire. The results of the pre-questionnaire showed that there were no significant differences in students' prior knowledge related to CT or the tools being used in the sessions. All the participating students had previous experience with Scratch; few (4%) had experience with Makey Makey; and almost 90% answered the CT problem correctly. Hence, prior knowledge was unlikely to be a confounding factor in the study.

5.2 Limitations

There are some limitations to our study. Firstly, due to the study design, it is not clear whether the demonstration of prototypes or DRP had a greater effect on the improved delayed test performance. Despite no new instruction or explicit review of theoretical concepts in the final sixth session, students showed their prototypes. It is possible that the showing and viewing of demos acted as a review of the CT concepts and affected student performance on the delayed test. Future studies should control for this possible confound. Secondly, the validity and reliability of the immediate and delayed tests should be established. These were the main instruments used to measure the learning and future studies should look to either validate the current tests or use a validated measure for assessing CT knowledge in primary school students such as the Computational Thinking Test (CTT) for middle school students by Bati and Kaan (Bati and Kaan, 2018). Thirdly, a control condition and a larger sample size would increase the reliability of the results. The control condition and sample size were affected by COVID-19 which prevented additional groups of students from completing the study. Additionally, it is unclear whether the enhanced retention of CT concepts persists for longer periods. Having students complete tests given at greater intervals would be beneficial. Finally, it is unclear whether these results can be generalized beyond the specific context of the study (i.e. with younger or older students).

5.3 Future Work

To overcome the limitations of the current study, a similar study could be conducted with a validated CT assessment, larger sample size, and an updated study design that holds the delayed test as an isolated activity occurring days or weeks after the last session, be it a learning or demonstration session, has taken place. Future work could also include a control condition, additional patterns for distributing the reviews, and tests given over longer periods of time in order to determine which DRP pattern performs better. Furthermore, alternatives to class-wide multi-choice tests using Kahoot! used for reviews could be explored.

6. CONCLUSION

To explore the effects of DRP on primary school student learning of a complex subject, in an authentic classroom setting, we conducted a quasi-experimental study involving 20 primary school students (ages 11-12). For the study, students participated in six weekly sessions on CT. To apply DRP, at the end of the first four classes, students took part in a class-wide Kahoot! review quiz that included material from previous classes. To assess student learning, a test was given at the end of the fifth session (immediate test) and another test was given a week later (delayed test) at the end of the demo class. The students were not warned about the tests to reduce the likelihood of students reviewing the materials prior to the tests. Although, students did perform 5-minute demos of their prototypes prior to the delayed test. Students attained higher scores on the delayed test in comparison to the immediate test. However, the difference in scores was not significant. The results suggest that DRP may play a role in helping students overcome the natural decay of memory, however, studies with a larger sample size are needed to examine if the proposed DRP scheme is effective.

In addition, due to limitations in the study design it is not clear whether the demonstration of prototypes or DRP had a greater effect on the improved delayed test performance. Overall, the study shows a simple implementation of DRP in an authentic classroom environment that involves primary school student learning of a complex subject such as CT. This simple implementation involves setting aside the final 15 minutes of a session for a class-wide multiple-choice review. To better support durable learning, further research involving simple, yet effective teaching practices conducted in authentic classroom settings is needed. Such research can both illustrate the benefits of using evidence-based practices and provide educators with an example of how such practices can be applied in their classrooms.

ACKNOWLEDGEMENT

We thank the school, teachers and students participating in this study. We also thank the Barcelona City Council and Barcelona Activa for trusting us to carry out this initiative. This work has been partially funded by the European Regional Development Fund, Erasmus+, and the National Research Agency of the Spanish Ministry of Science, Innovation and Universities under project grants TIN2017-85179-C3-3-R, 2017-1-ES01-KA201-038220 and MDM-2015-0502. Dr. Hernández-Leo acknowledges the support by ICREA under the ICREA Academia programme.

REFERENCES

Bati, K., 2018. Computational Thinking Test (CTT) for Middle School Students. Mediterranean Journal of Educational Research, Vol. 12. pp. 89-101. doi: 10.29329/mjer.2018.138.6.

Bocconi, S. et al. 2016. Exploring the Field of computational Thinking as a 21st Century Skill, EDULEARN16 Proceedings, pp. 4725-4733.

Butler, A. C. and Roediger, H. L., 2008. Feedback enhances the positive effects and reduces the negative effects of multiple-choice testing, Memory and Cognition, Vol. 36, No.3, pp. 604–616. doi: 10.3758/MC.36.3.604.

Cecilio-Fernandes, D. et al, 2017. Comparison of formula and number-right scoring in undergraduate medical training: A rasch model analysis. BMC Medical Education Vol. 17, pp. 192.

Davis, R. L. and Zhong, Y., 2017. The Biology of Forgetting—A Perspective, Neuron. Elsevier Inc., Vol. 95, No. 3, pp. 490–503. doi: 10.1016/j.neuron.2017.05.039.

Denning, P. J., 2017. Remaining trouble spots with computational thinking. Communications of the ACM, Vol. 60 No. 6, pp.33-39.

Dobson, J. L., et al, 2017. Distributed retrieval practice promotes superior recall of anatomy information, Anatomical Sciences Education. John Wiley and Sons Inc., Vol. 10 No. 4, pp. 339–347. doi: 10.1002/ase.1668.

European Commission (2019) The Computational Thinking Study | EU Science Hub, EU Science Hub. Available at: https://ec.europa.eu/jrc/en/computational-thinking (Accessed: 24 July 2020).

Halpern, D. F. and Hakel, M. D., 2002. Learning That Lasts a Lifetime: Teaching for Long-Term Retention and Transfer. New Directions for Teaching and Learning. John Wiley & Sons, Ltd, Vol. 89, pp. 3–7. doi: 10.1002/tl.42.

Kelleher, C. and Pausch, R., 2005. Lowering the Barriers to Programming: a survey of programming environments and languages for novice programmers. ACM Computing Surveys - CSUR.

Muijtjens, A. et al, 1999. The effect of a "don't know" option on test scores: number-right and formula scoring compared. Medical Education, pp.267–275.

Murray, S. R. and Udermann, B. E., 2003. Massed versus Distributed Practice: Which is Better?, CAHPERD Journal. Vol. 28, No. 1.

Murre, J.M. and Dros, J., 2015. Replication and analysis of Ebbinghaus' forgetting curve. PloS one, Vol. 10, No. 7.

Nakata, T. et al, 2020. Effects of Distributed Retrieval Practice Over a Semester: Cumulative Tests as a Way to Facilitate Second Language Vocabulary Learning, TESOL Quarterly. Wiley Blackwell, p.596. doi: 10.1002/tesq.596.

Papavlasopoulou, S., et al, 2017. Empirical studies on the Maker Movement, a promising approach to learning: A literature review.' Entertainment Computing, Vol.18, pp. 57-78.

Pashler, H. et al, 2005. When Does Feedback Facilitate Learning of Words? Journal of Experimental Psychology: Learning, Memory, and Cognition, Vol 31, No. 1, pp. 3-8

Pazzani, M. J., 1991. Influence of prior knowledge on concept acquisition: Experimental and computational results. Journal of Experimental Psychology: Learning, Memory, and Cognition, Vol. 17 No.3, pp. 416–432 doi:10.1037/0278-7393.17.3.416.

Rawson, K. A. and Kintsch, W., 2005. Rereading effects depend on time of test, Journal of Educational Psychology, Vol. 97 No. 1, pp. 70–80. doi: 10.1037/0022-0663.97.1.70.

Resnick, M. et al, 2009 Scratch: Programming for All, Communications of the ACM, Vol. 5 No. 11. doi: 10.1145/1592761.1592779.

Rich, P. J., et al, 2013 Convergent cognition, Instructional Science. Springer, Vol. 41, No. 2, pp. 431–453. doi: 10.1007/s11251-012-9240-7.

Roediger III, H.L. and Karpicke, J.D., 2006. Test-enhanced learning: Taking memory tests improves long-term retention. Psychological Science, Vol. 17, No. 3, pp. 249-255.

Rohrer, D., 2009. The Effects of Spacing and Mixing Practice Problems, Journal for Research in Mathematics Education, Vol. 40, No. 1, pp. 4–17. doi: 10.2307/40539318.

Silver J, et al., 2012. Makey Makey improvising tangible and nature-based user interfaces. In Proceedings of the ACM tangible embedded and embodied interaction, Kingston, Ontario, Canada, pp 367–370.

Wang, A. I. and Tahir, R., 2020. The effect of using Kahoot! for learning – A literature review, Computers and Education. Elsevier Ltd, Vol. 149, p. 103818. doi: 10.1016/j.compedu.2020.103818..

Yadav, A. et al, 2016. Computational Thinking for All: Pedagogical Approaches to Embedding 21st Century Problem Solving in K-12 Classrooms. TechTrends Vol. 60, pp. 565–568doi: 10.1007/s11528-016-0087-7.